CHAPTER 2

GENIC MOLECULAR MARKERS IN PLANTS: DEVELOPMENT AND APPLICATIONS

RAJEEV K. VARSHNEY[1,*], THUDI MAHENDAR[1],
RAMESH K. AGGARWAL[2] AND ANDREAS BÖRNER[3]

[1]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru - 502324, India*

[2]*Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad- 500 007, India*

[3]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Correnstrasse 3, D-06466 Gatersleben, Germany*

Abstract: The current advancement in plant biology research encompassing: generation of huge amount of molecular-genetic data, development of impressive methodological skills in molecular biology experimentation, and systems analyses, has set the stage to search for ways/means to utilize the available resources to strengthen interdisciplinary efforts to find solutions to the challenging goals of plant breeding efforts (such as abiotic stress tolerance) ultimately leading to gainful applications in crop improvement. A positive fall out of such a realization and efforts has been the identification/development of a new class of very useful DNA markers called genic molecular markers (GMMs) utilizing the ever-increasing archives of gene sequence information being accumulated under the EST sequencing projects on a large number of plant species in the recent years. These markers being part of the cDNA/EST-sequences, are expected to represent the functional component of the genome i.e., gene(s), in contrast to all other random DNA-based markers (RDMs) that are developed/generated from the anonymous genomic DNA sequences/domains irrespective of their genic content/information. Therefore, identifying DNA sequences that demonstrate large effects on adaptive plant behavior remains fundamental to the development of GMMs. The few recent studies have now demonstrated the utility of these markers in genetic studies, and also shown that GMMs may be superior than RDMs for use in the marker-assisted selection, comparative mapping, and exploration of the functional genetic diversity in the germplasm adapted to different environments. The only constraint of GMMs is their low level of polymorphism as compared to the RDMs, which is expected of their origin from the relatively conserved functional portion of the genome. This chapter provides a critical review of the development and various applications of the GMMs.

*Corresponding Authors: r.k.varshney@cgiar.org

R.K. Varshney and R. Tuberosa (eds.), Genomics-Assisted Crop Improvement: Vol. 1: Genomics Approaches and Platforms, 13–29.

1. MOLECULAR MARKERS IN PLANT BREEDING

In agriculture, one of the main objectives of plant breeder is to improve the existing cultivars, which are deficient in one or more traits by crossing such cultivars with lines that possess the desired trait. A conventional breeding programme thus involves crossing whole genomes followed by selection of the superior recombinants from among the several segregation products. Indeed, such a procedure is laborious and time consuming, involving several crosses, several generations, and careful phenotypic selection, and the linkage drag (tight linkage of the undesired loci with the desired loci) may make it further difficult to achieve the desired objective. Advent of DNA marker technology, development of several types of molecular markers and molecular breeding strategies offered possibilities to plant breeders and geneticists to overcome many of the problems faced during conventional breeding.

Molecular markers are now widely used to track loci and genome regions in several crop-breeding programmes, as molecular markers tightly linked with a large number of agronomic and disease resistance traits are available in major crop species (Phillips and Vasil 2001, Jain et al. 2002, Gupta and Varshney 2004). These molecular markers include: (i) hybridization-based markers such as restriction fragment length polymorphism (RFLP), (ii) PCR-based markers: random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellite or simple sequence repeat (SSR), and (iii) sequence-based markers: single nucleotide polymorphism (SNP). The majority of these molecular markers has been developed either from genomic DNA libraries (e.g. RFLPs and SSRs) or from random PCR amplification of genomic DNA (e.g. RAPDs) or both (e.g. AFLPs). These DNA markers can be generated in large numbers and can prove to be very useful for a variety of purposes relevant to crop improvement. For instance, these markers have been utilized extensively for the preparation of saturated molecular maps (genetical and physical). Their association with genes/QTLs controlling the traits of economic importance has also been utilized in some cases for indirect marker-assisted selection (MAS) (e.g. Koebner 2004, Korzun 2002). Other uses of molecular markers include gene introgression through backcrossing, germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organization and phylogenetic analysis (see Jain et al. 2002). For plant breeding applications, SSR markers, among different classes of the existing markers, have been proven and recommended as markers of choice (Gupta and Varshney 2000). RFLP is not readily adapted to high sample throughput and RAPD assays are not sufficiently reproducible or transferable between laboratories. While both SSRs and AFLPs are efficient in identifying polymorphisms, SSRs are more readily automated (Shariflou et al. 2001). Although AFLPs can in principle be converted into simple PCR assays (e.g. STSs), this conversion can become cumbersome and complicated as individual bands are often composed of multiple fragments (Shan et al. 1999), particularly in large genome templates.

2. GENIC MOLECULAR MARKERS: INTRODUCTION AND DEVELOPMENTS

Due to emphasis on functional genomics, several gene discovery projects in the form of genome sequencing, transcriptome sequencing or gene expression studies have been established since last five years. As a result, a large number of genes have been identified through 'wet lab' as well as *in silico* studies and a wealth of sequence data have been accumulated in public databases (e.g. http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk) in the form of BAC (bacterial artificial chromosome) clones, ESTs (expressed sequence tags), full length cDNA clones and genes. The availability of enormous amount of sequence data from complete or partial genes has made it possible to develop the molecular markers directly from the parts of genes. These markers are referred as "genic" molecular markers (GMM).

The majority of the markers, developed and used in the past as described above in section 1, are directly derived from the genomic DNA, and therefore could belong to either the transcribed or the non-transcribed part of the genome without any information available on their functions. In contrast, GMMs developed from coding sequences like ESTs or fully characterized genes frequently have been assigned known functions. Based on the site of polymorphism and later's effect on phenotypic variation, GMMs have been classified into two groups (Anderson and Luebberstedt 2003):

(i) Gene-targeted markers (GTMs): derived from polymorphisms within genes, however not necessarily involved in phenotypic trait variation, e.g. untranslated regions (UTRs) of EST sequences (Schmitt et al. 2006; Aggarwal et al 2007);

(ii) Functional markers (FMs): derived from polymorphic sequences or sites within genes and, thus, more likely to be causally involved in phenotypic trait variation (e.g. candidate gene-based molecular markers). The FMs, depending on the involvement in the phenotypic trait variation, are further classified into two subgroups: (a) indirect functional markers (IFMs), for which the role for phenotypic trait variation is indirectly known, and (b) direct functional markers (DFMs), for which the role for the phenotypic trait variation is well proven.

As per the above terminology, the molecular markers derived from anonymous regions of the genome are called random DNA markers (RDMs), which may or may not be developed from the polymorphic site in gene or may not be developed from a gene at all.

Although genic markers were developed earlier also, these were in the form of cDNA–RFLP (Graner et al. 1991, Causse et al. 1994) for which functions could not be predicted at that time. However, some efforts were made to sequence these early cDNA clones to determine the genes and their functions (Michalek et al. 1999). Compared to these earlier efforts, development of genic markers have become a reality only in recent years, because of accumulation of large ESTs or gene sequences resources resulting from EST and genome sequencing projects in several crop species and also due to the developments in the field of bioinformatics (Gupta and Rustgi 2004). For example, several transcriptome resources have

become available (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), and software tools or pearl scripts have been developed to search for SSRs and SNPs from EST or gene sequences (Varshney et al. 2004, 2005a).

Although, whole genome sequencing and annotation is the way to identify the entire gene repository of a species, this has been possible only for a limited number of crop species involving large scale sequencing of their genome or gene space. On the other hand, ESTs represent a basic commodity within the analysis of genomes and their genes for a species (Rudd et al. 2003). Whereas the complete sequencing of a genome may utilize either a clone-by-clone approach or a whole genome shotgun approach to acquire adequate coverage to assemble a meaningful scaffold, EST sequencing is directed at the quick, cheap and simple sequencing of partial gene transcripts (Sreenivasulu et al. 2002). As a result, a significant redundancy can be observed in gene sequence data obtained from EST sequencing projects (see Varshney et al. 2004). Therefore before developing molecular markers from ESTs, it is essential to define the "unigenes" after cluster analysis of random ESTs using appropriate computer programmes such as stackPack (Miller et al. 1999).

Once the unigene sequence data from EST analysis or non-redundant set of genes are available, molecular markers can be developed using two main approaches:

(1) Direct mapping: Under this approach, either the cDNA clones corresponding to the ESTs of interest can be used as RFLP probe or the PCR primers can be

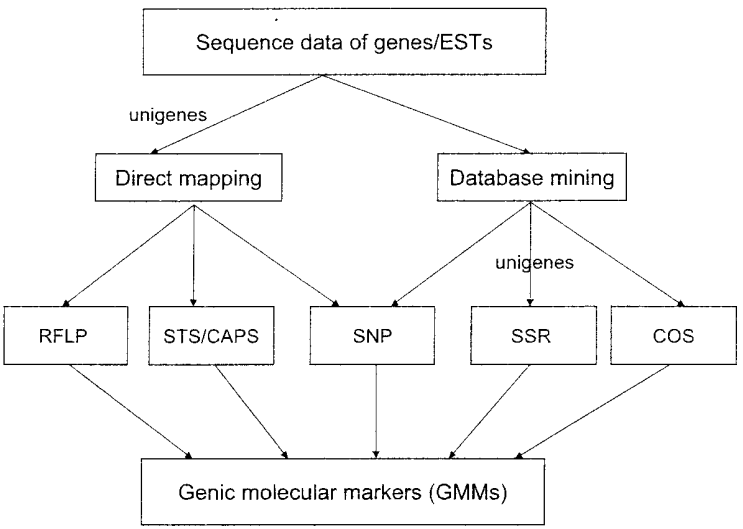

Figure 1. A scheme for development of genic molecular markers (GMMs). Two common ways to develop GMMs are shown in the figure. In the first method, the sequence data are used to define the unigenes and then the cDNA clones or genic clones corresponding to the unigenes can be assayed as RFLPs or the unigene sequence data can be used to design the primer pairs and assayed using STS/CAPS or SNP assays. In the second method, the sequence data can be mined by using some computer programmes or scripts to identify the SSRs, SNPs or COSs from given sequence data and then these markers, after defining the unigenes, can be assayed using appropriate genotyping platforms

designed for the EST/gene and used as STS or CAPS marker. Direct mapping approach should be undertaken with the unigene set of ESTs or genes only.

(2) *In silico* mining: In this approach, the SSR or SNP identification software tools are used to screen the sequence data for ESTs/genes. For identification of SNPs, the redundant set of EST data, generated from more than one genotype of a given species, are used. However, after identification of SNPs, only non-redundant set of ESTs should be considered for SNP mapping.

A scheme for development of GMMs has been shown in Figure 1. Development of FMs, however, requires: (i) functionally characterized genes, (ii) allele sequences from such genes, (iii) identification of polymorphic, functional motifs affecting plant phenotype within these genes, and (iv) validation of associations between DNA polymorphisms and trait variation. Therefore depending on the objective as well as available information or feasibility, the FMs, the special class of GMMs, can also be generated.

3. APPLICATIONS OF GENIC MOLECULAR MARKERS

Molecular markers have already shown their applications in a variety of ways in several plant species (see Gupta and Varshney 2004). The development of GMMs, now permits a targeted approach for detection of nucleotide diversity in genes controlling agronomic traits in plant populations. Some main areas of plant breeding and genetics, where the implementation of GMMs will prove quite useful, are discussed here.

3.1. Trait Mapping

One of the main applications of molecular markers in plant breeding is their use as diagnostic markers for the trait in the selection. However, use of random molecular markers (RDMs) as a diagnostic tool entails the risk of losing the linkage through genetic recombination. Even in case of GMMs, the gene-targeted markers (GTMs) where polymorphism was discovered through one allele analysis without any further specification of the polymorphic sequence motif are threatened by the same way (Rafalski and Tingey, 1993). In contrast to RDMs or GTMs, FMs (DFMs or IFMs) allow reliable application of markers in populations without prior mapping and the use of markers in mapped populations without risk of information loss owing to recombination.

The development of FMs is expensive and cannot be undertaken for all the traits and in all crop species, GMM have been developed and mapped in several plant species (Table 1). The genetic maps, developed after mapping/integration of GMM are called "transcript" or "gene" maps. For example, based on the candidate genes for drought tolerance, a comprehensive set of >200 gene-based markers have been developed for barley (Rostocks et al. 2005). Recently, a "transcript map" of barley after integrating more than 1000 gene-based markers (GTMs) has been developed, (Stein et al. 2007). A kind of transcriptome map based on deletion mapping of

Table 1. Some reports on development of genic molecular markers in important plant species

General name	Species	Type of markers developed	References
Cereals and grasses			
Barley	*Hordeum vulgare*	EST-SSR, EST-SNP, EST-RFLP, cDNA-RFLP	Thiel et al. 2003, Rostocks et al. 2005, Varshney et al. 2006, Willsmore et al. 2006, Stein et al. 2007, Varshney et al. 2007b
Maize	*Zea mays*	cDNA-RFLP, EST-SNP	Gardiner et al. 1993, Chao et al. 1994, Picoult-Newberg et al. 1999, Falque et al. 2005
Wheat	*Triticum aestivum*	EST-SSR, EST-SNP, cDNA-RFLP	Holton et al. 2002, Yu et al. 2004, Somers et al. 2003, Gao et al. 2004, Qi X. et al. 2004, Nicot et al. 2004
Rice	*Oriza sativa*	EST-SSR, EST-SNP, cDNA-RFLP, Intron Length Polymorphism (ILP)	Causse et al. 1994, Harushima et al. 1998, Temnykh et al. 2001, Feltus et al. 2004, Wang et al. 2005
Rye	*Secale cereale*	EST-SSR, EST-SNP	Hackauf and Wehling, 2002, Khlestkina et al. 2004, Varshney et al. 2007b
Sorghum	*Sorghum bicolor*	EST-SSR, cDNA-RFLP	Childs et al. 2001, Klein et al. 2003, Bowers et al. 2003, Ramu et al. 2006, Jayashree et al. 2006
Lolium	*Lolium perenne*	EST-SSR	Faville et al. 2004
Legumes			
White clover	*Trifolium repens*	EST-SSR	Barret et al. 2004
Soybean	*Glycine max*	EST-SSR	Song et al. 2004, Zhang et al. 2004
Fiber and oil seed crops			
Cotton	*Gossypium* sps.	EST-SSR	Zhang et al. 2005, Chee et al. 2004, Park et al. 2005-
Sunflower	*Helianthus* sps.	EST-SNP	Lai et al. 2005
Fruit and vegetables			
Grape	*Vitis vinifera*	EST-SSR	Chen et al. 2006
Kiwi fruit	*Actinidia chinensis*	EST-SSR	Fraser et al. 2004
Raspberry	*Rubus* spp.	EST-SSR	Graham et al. 2004
Tomato	*Lycopersicon esculentum*	EST-SSR	Frary et al. 2005
Strawberry	*Fragaria* spp.	EST-SSR	Sargent et al. 2006
Trees			
Pinus	*Pinus* ssp.	EST-SSR, ESTP	Cato et al. 2001
Coffee	*Coffea* ssp.	EST-SSR	Bhat et al. 2005, Aggarwal et al. 2007

more than 16,000 gene loci has been developed in wheat (Qi L-L et al. 2004). Such molecular maps, not only provide gene based molecular markers associated with the trait of interest after the QTL analysis, but also can be compared with those of the other related plant species in an efficient manner.

3.2. Functional Diversity

Characterization of genetic variation within natural populations and among breeding lines is crucial for effective conservation and exploitation of genetic resources for crop improvement programmes. Molecular markers have proven useful for assessment of genetic variation in germplasm collections (Hausmann et al. 2004; Maccaferri et al. 2006). Evaluation of germplam with GMMs might enhance the role of genetic markers by assaying the variation in transcribed and known function genes, although there may be a higher probability of bias owing to selection.

While using the genic SSR markers for diversity studies, the expansion and contraction of SSR repeats in genes of known function can be tested for association with phenotypic variation or, more desirably, biological function (Ayers et al. 1997). The presence of SSRs in the transcripts of genes suggests that they might have a role in gene expression or function; however, it is yet to be determined whether any unusual phenotypic variation might be associated with the length of SSRs in coding regions as was reported for several diseases in human (Cummings and Zoghbi 2000). Similarly, the use of SNP markers for diversity studies may correlate the SNPs of coding *vs.* non-coding regions of the gene with the trait variation. The variation associated with deleterious characters, however, is less likely to be represented in the germplasm collections of crop species than among natural populations because undesirable mutations are commonly culled from breeding populations (Cho et al. 2000).

Several studies involving GMMs, especially genic SSRs, have been found useful for estimating genetic relationship on one hand (see Gupta et al. 2003 Gupta and Rustgi 2004, Varshney et al. 2005a) while at the same time these have provided opportunities to examine functional diversity in relation to adaptive variation (Eujayl et al. 2001, Russell et al. 2004). It seems likely that with the development of more GMMs in major crop species, genetic diversity studies will become more meaningful by a shift in emphasis from the evaluation of anonymous diversity to functional genetic diversity in the near future. Nevertheless, use of the neutral RDM markers will remain useful in situations where: (i) GMMs would not be available, and (ii) to address some specific objectives e.g. neutral grouping of germplasm.

3.3. Interspecific or Intergeneric Transferability

Perhaps one of the most important features of the GMMs is that these markers provide high degree of transferability among distantly related species. In contrast, except RFLPs all other RDMs are generally constrained in this regard. Transferability of GMM markers to related species or genera has now been demonstrated in several studies (Table 2). For example, a computational study based on analysis

Table 2. Some examples of interspecific or intergeneric transferability of genic molecular markers

Plant species	Marker type	Species, recorded transferability	Reference
Cereals and grasses			
Barley (*Hordeum vulgare*)	EST-SSR, EST-SNP	Wheat, rice, rye	Thiel et al. 2003, Varshney et al. 2004, 2007b
Wheat (*Triticum aestivum*)	EST-SSR	*Aegilops* and *Triticum* species, barley, maize, rice, rye, oats, soybean, *Lophopyrum elongatum*	Holton et al. 2002, Gupta et al. 2003, Gao et al. 2003, Bandopadhyay et al. 2004, Yu et al. 2004, Mullan et al. 2005, Tang et al. 2006
Rice (*Oryza sativa*)	EST-SSR	wild species of rice	Cho et al. 2000
Sugarcane (*Saccharaum officinarum*)	EST-SSR	*Saccharum robustum*, *Erianthus* and Sorghum	Cordeiro et al. 2001
Sorghum (*Sorghum bicolor*)	EST-SSR	*Eleusine coracana*, *Seashore paspalum*, finger millet	Wang et al. 2005
Tall fescue (*Festuca*)	EST-SSR	subfamilies of Poaceae	Mian et al. 2005
Fiber and oilseed crops			
Cotton (*Gossypium hirsutum*)	EST-SSR	Cotton species	Saha et al. 2003
Sunflower (*Helianthus annus*)	EST-SSR	*Heliantus angustifolius*, *Helianthus verticillatus*	Pashley et al. 2006
Fruit and vegetables			
Strawberry (*Fragaria vesca*)	EST-SSR	*F. gracilis*, *F. iinumae*, *F. nilgerrensis*, *F. nipponica*	Bassil et al. 2006
Apricot (*Prunus armeniaca*)	EST-SSR	Vitaceae and Roseaceae family	Decroocq et al. 2003
Grape (*Vitis vinifera*)	EST-SSR	> 25 species from 5 Vitaceae and Roseaceae	Scott et al. 2000, Rossetto et al. 2002, Arnold et al. 2002, Decroocq et al. 2003
Tomato (*Solanum lycopersicum*)	EST-SSR	Solanaceous members	Frary et al. 2005
Ferns and trees			
Alpine lady-fern (*Atyrium distentifolium*)	EST-SSR	9 species from Woodsiaceae	Woodhead et al. 2003
Pinus (*Pinus taeda*)	EST-SSR	12 *Pinus* species	Komulainen et al. 2003, Changne et al. 2004, Liewlaksaneeyanawin et al. 2004
Spruce (*Picea glauca*)	EST-SSR	23 *Picea* species	Rungis et al. 2004
Citrus (*Citrus sinensis*)	EST-SSR	*Poncirus trifoliata*	Chen et al. 2006
Coffee (*Coffea arabica*, *Coffea canephora*)	EST-SSR	16 species of coffee and *Psilanthus*	Bhat et al. 2005, Poncet et al. 2006, Aggarwal et al. 2007

of ~1000 barley GMMs suggested a theoretical transferability of barley markers to wheat (95.2%), rice (70.3%), maize (69.3%), sorghum (65.9%), rye (38.1%) and even to dicot species (~16%). Infact, *in silico* analyses of GMMs of wheat, maize and sorghum with complete rice genome sequence data have provided a larger number of anchoring points among different cereal genomes as well as provided insights into cereal genome evolution (Sorrells et al. 2003, Salse et al. 2004).

In some studies, the use of GMMs of major crop species has been shown to enrich the genetic maps of related plant species for which little marker information is available. For example, barley EST-SSR as well as EST-SNP markers have been shown transferable as well as mappable in syntenic regions of rye (Varshney et al. 2004, 2005c, 2007a; Figure 2). Further, such kind of markers from the related plant species offers the possibility to develop anchor or conserved orthologous sets (COS) for genetic analysis and breeding in different species. In this direction, Rudd et al. (2005) identified a large repository of such COS markers and developed a database called "PlantMarker".

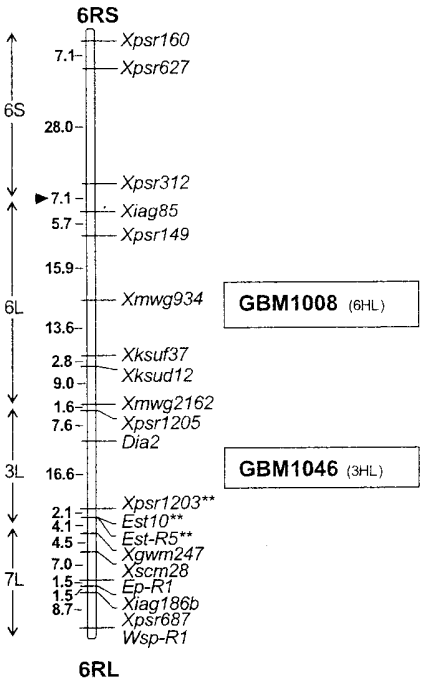

Figure 2. An example of integration of barley genic (EST-SSR) markers into syntenic regions of rye genetic map. Integrated barley markers (GBM1008, GBM1046) are shown in bold and capital font in boxes on right hand side. Details about other markers present on this linkage group are available in Korzun et al. (2001). Genetic distances are given in centimorgans (cM) on left hand side. The black triangle indicates the estimated centromere position. The relationship of the linkage group 6R in terms of Triticeae linkage group is shown on very left hand side (left to black triangle) as per Devos et al. (1993). Both barley genic markers from linkage group 3H and 6H are mapped into expected syntenic regions of the rye linkage group 6R. S = short arm, L = long arm

4. COMPARISON OF GMMs AND RDMs

Since the development of first molecular markers i.e. RFLPs in 1980 (Botstein et al. 1980), a diverse array of molecular marker technologies have come into being revolutionizing conventional plant breeding efforts for crop improvement. Significant strides have been made in crop improvement through conventional random molecular markers (RDMs). For instance, these molecular markers besides throwing light on organization, conservation and evolution of plant genomes, have also aided geneticists and plant breeders to tag genes, map QTLs for the traits of economic importance. Still, most of them are "anonymous" markers, that is to say their biological function is unknown. In comparison, a putative function for majority of the molecular markers, derived from the genes or ESTs, however can be deduced using some bioinformatics tools; such markers (GMMs) are commonly referred as functional markers (Varshney et al. 2005b). Although, in *stricto* sense, the functional markers are based on functionally defined genes underlying specific biochemical or physiological functions and therefore the FMs can be considered as a class of GMMs (Anderson and Luebberstedt 2003).

The GMMs, like RDMs, could detect both length and sequence polymorphisms in expressed regions of the genome but provide relatively stronger and robust marker assays. However, as compared to the RDMs the developmental costs of GMMs, depend on which specific class of GMMs is to be developed. Similarly the applied value of the GMMs as compared to the RDMs varies depending on the class of the GMMs. These relative costs and applications issues have been detailed in Table 3. In summary, if the GMMs based on the polymorphic site and verification are developed (i.e. FMs), these markers are superior to RDMs for using them as diagnostic tools in marker-assisted selection as they may owe the complete linkage with the trait locus alleles (Anderson and Luebberstedt 2003). In plant breeding, the GMMs are superior to RDMs for selection of, e.g., parent materials to build segregating populations, as well as subsequent selection of lines (line breeding) or inbreds (hybrid breeding). Depending on the mode of the GMM characterization, these can also be applied to the targeted combination of alleles in hybrid and synthetic breeding. In population breeding and recurrent selection programs, the GMMs can be employed to avoid genetic drift at characterized loci.

Being originated from the conserved proportion of the genome, the GMMs, as compared to the RDMs, are the candidate markers for interspecific/intergeneric transferability and comparative mapping/genomics studies in related plant species. Since the GMMs represent the expressed portion of the genome, they sample the variation in transcribed regions of the genome, and provide a more direct estimate of functional diversity while screening the markers on the germplasm adapted to different environments. Nevertheless, the GMMs, as compared to the RDMs are less polymorphic and provide less alleles and lower PIC values. Additionally, due to biased distribution in the genome, the GMMs are unsuitable for analyzing population structure.

Table 3. Comparison of genic molecular markers (GMMs) with random DNA markers (RDMs)

Feature	GMMs	RDMs		
		gSSRs, SNPs	RFLPs	RAPD/AFLP/ ISSR etc.
Need for sequence data	Genes/ESTs data Essential	Essential	Not required	Not required
Costs of generation	Low*	High	High	Low-moderate
Labour involved	Less	Much	Much	Less
Level of polymorphism	Low	High	Low	Low-moderate
Interspecific transferability and comparative mapping	High	Low-moderate	Moderate-High	Low-moderate
Function of markers	Known majority of times	Unknown majority of times	Unknown	Unknown
Utility in marker-assisted selection	Great, if the marker is derived from the gene, involved in expression of trait	High	Moderate	Low-moderate

*generally GMMs are by products of the available transcriptome resources being developed for functional genomic studies.

5. FUTURE DIRECTIONS OF GENIC MOLECULAR MARKERS

It is clear that the GMMs and especially the FMs are extremely useful source of markers in plant breeding for marker-assisted selection because these markers may represent the genes responsible for expression of target traits. If so, there will not be any recombination between the markers and the trait, thus representing perfect indirect selection tools. While low level of polymorphism is an inherent feature of the GMMs, it is compensated by their higher interspecific transferability as well as capacity to sample the functional diversity in the germplasm. These features make the development and application of the GMMs more attractive for plant breeding and genetics.

With more DNA sequence data being generated continuously, the trend is towards cross-referencing genes and genomes using sequence and map-based tools. Because polymorphism is a major limitation for many species, SSR- and SNP-based GMMs will be valuable tools for plant geneticists and breeders. In the longer term, development of allele-specific, functional markers (FMs) for the genes controlling agronomic traits will be important for advancing the science of plant breeding. In this context genic SSR and SNP markers together with other types of markers that target functional polymorphisms within genes will be developed in near future for major crop species. The choice of the most appropriate

marker system, however, needs to be decided on a case-by-case basis and will depend on many issues including the availability of technology platforms, costs for marker development, species transferability, information content and ease of documentation.

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishna KV, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Arnold C, Rossetto M, McNally J, Henry RJ (2002) The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in *Vitaceae*. Am J Bot 89:22–28

Ayers NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Bandopadhyay R, Sharma S, Rustgi S, Singh R, Kumar A, Balyan HS, Gupta PK (2004) DNA polymorphism among 18 species of *Triticum–Aegilops* complex using wheat EST-SSRs. Plant Sci 166:349–356

Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D (2004) A microsatellite map of white clover. Theor Appl Genet 109:596–608

Bassil NV, Njuguna W, Slovin JP, (2006) EST-SSR markers from *Fragaria vesca* L. cv. yellow wonder. Mol Ecol Notes 6:806–809

Bhat PR, Kumar KV, Hendre PS, Kumar RP, Varshney RK, Aggarwal RK (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeats, markers from robusta coffee variety 'C × R' (an interspecific hybrid of *Coffea canephora* × *Coffea congensis*). Mol Ecol Notes 5:80–83

Botstein D, White RI, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z (2003) A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR based method for genetically mapping ESTs. Theor Appl Genet 102:296–306

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274

Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109:1204–1214

Chao S, Baysdorfer C, Heredia-Diaz O, Musket T, Xu G, Coe Jr. EH (1994) RFLP mapping of partially sequenced leaf cDNA clones in maize. Theor Appl Genet 88:717–721

Chee PW, Rong JK, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers homologues in cotton. Genome 47:449–462

Chen C, Zhou P, YA, Huang S, Gmitter Jr. FG (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Childs KL, Klein RR, Klein PE, Morishige DT, Mullet JE (2001) Mapping genes on an integrated sorghum genetic and physical map using cDNA selection technology. Plant J 27:243–256

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Parl WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro GM, Casu R, McIntyre L, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to *Erianthus* and *Sorghum*. Plant Sci 160:1115–1123

Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. Hum Mol Genet 9:909–916

Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912–922

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD (1993) Chromosomal rearrangement in the rye genome relative to that of wheat. Theor Appl Genet 85:673–680

Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39–43

Falque M, Décousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribière N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene 1 sequences or sites with in genes loci. Genetics 170:1957–1966

Faville MJ, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster W (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:12–32

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson (2004) An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. Genetics 148:479–494

Frary A, Xu Y, Liu J, Tedeschi SME, Tanksley S (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theor Appl Genet 111:291–312

Fraser LG, Harvey CF, Crowhurst RN, De Silva HN (2004) EST-derived microstellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108:1010–1016

Gao L, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gardiner J, Coe MEH, Melia-Hancock S, Hoisington DA, Chao S, (1993) Development of a core RFLP map in maize using an immortalized F_2 population. Genetics 134:917–930

Graham J, Smith K, Mac Kenzie K, Jorgenson L, Hackett C, Powell W (2004) The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. *idaeus*) based on AFLPs, genomic-SSR and EST-SSR markers. Theor Appl Genet 109:740–749

Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an RFLP map of barley. Theor Appl Genet 83:250–256

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetics and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Varshney RK (2004) Cereal genomics: an overview In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Press, Dordrecht, The Netherlands, p 639

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 121:17–25

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A (1998) A high-density rice genetic linkage map with 2275 Markers using a single F_2 population. Genetics 148:479–494

Haussmann BI, Hess DE, Omanya GO, Folkertsma RT, Reddy BV, Kayentao M, Welz HG, Geiger HH, (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016

Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9:63–71

Jain SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer Academic Publishers, The Netherlands

Jayashree B, Ramu P, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. In Silico Biol 6:0054

Khlestkina EK, Than MHM, Pestsova EG, Roder MS, Malyshev SV, Korzun V, Borner A (2004) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 109:725–732

Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J 34:605–622

Koebner RMD (2004) Marker-assisted selection in the cereals: the dream and the reality. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, Netherlands, p 199

Komulainen P, Brown GR, Mikkonen M, Karhu A, Garcia-Gil MR, Malley DO, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. Theor Appl Genet 107:667–678

Korzun V (2002) Use of molecular markers in cereal breeding. Cell Mol Biol Lett 7:811–820

Korzun V, Malyshev S, Voylokov AV, Börner A (2001) A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, protein, microsatellite and gene loci. Theor Appl Genet 102:709–717

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Maccaferri M, Sanguineti MC, Natoli E, Araus-Ortega JL, Ben Salem M, Bort J, Chenenaoui S, Deambrogio E, Garcia DML, De Montis A et al (2006) A panel of elite accessions of durum wheat (*Triticum durum* Desf.) Suitable for association mapping studies. Plant Genet Res 4:79–85

Mian MAR, Saha MC, Hopkins AA, Wang ZY (2005) Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. Genome 48:637–647

Michalek W, Kunzel G, Graner A (1999) Sequence analysis and gene identification in a set of mapped RFLP markers in barley (*Hordeum vulgare*). Genome 42:849–853

Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide WA (1999) A comprehensive approach to clustering of expressed human gene Sequence: the sequence tag alignment and consensus knowledge base. Genome Res 9:1143–1155

Mullan DJ, Platteter A, Teakle NL, Appels R, Colmer TD, Anderson J, Francki M (2005) EST-derived SSR markers from defined regions of the wheat genome to identify *Lophopyrum elongatum* specific loci. Genome 48:811–822

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428–441

Pashley CH, Ellis JR, Mccauley DE, Burke JM (2006) EST Databases as a source for molecular markers: lessons from helianthus. J Hered 97:381–388

Phillips RL, Vasil IK (2001) DNA-based markers in plants. In: Phillips RL, Vasil IK (eds) DNA-based markers in plants. Kluwer Academic Publishers, Dordrecht, The Netherlands, p 497

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167–174

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST–SSRs as markers for the *Coffea* genus. Mol Genet Genomics 276:436–449

Qi L-L, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri et al (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

Qi X, Pittaway TS, Lindup TS, Liu S, Liu H, Waterman E, Padi FK, Hash CT, Zhu J, Gale, MD, Devos KM (2004) An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, *Pennisetum glaucum*. Theor Appl Genet 109:1485–1493

Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet 9:275–280

Ramu P, Bantte K, Ashok CK, Jayashree B, Rolf TF, Senthilvel S, Mahalakshmi V, Reddy AL, Fakrudin B, Hash CT (2006) Development and mapping of EST-SSR markers in sorghum for comparative mapping with rice. In: Plant and animal genome XIV conference, San Diego, CA, USA, P 200 (http://www.intl-pag.org/14/abstracts/PAG14_P200.html)

Rossetto M, McNally J, Henry RJ (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the *Vitaceae*. Theor Appl Genet 104:61–66

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson J, Wanamaker S, Walia H, Rodriguez E, Hedley P, Liu H, Morris J, Close T, Marshall D, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics 274:515–527

Rudd S, Mewes HW, Mayer KF (2003) Sputnik: a database platform for comparative plant genomics. Nucleic Acids Res 31:128–132

Rudd S, Schoof H, Klaus M (2005) PlantMarkers-A database of predicted molecular markers from plants. Nucleic Acids Res 33:D628–D632

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet 109:1283–1294

Russell J, Booth A, Fuller J, Harrower B, Hedley P (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Salse J, Piegu B, Cooke R, Delseny M (2004) New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) Highlights reshuffling and identifies new duplications in the rice genome. Plant J 38:396–409

Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Aru´s P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH (2006) An enhanced microsatellite map of diploid *Fragaria*. Theor Appl Genet 112:1349–1359

Schmitt BA, Costa JH, Dirce Fernandes de Melo (2006) AOX –A functional marker for efficient cell reprogramming under stress? Trends Plant Sci 11:281–287

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Shan X, Blake TK, Talbert LE (1999) Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. Theor Appl Genet 98:1072–1078

Shariflou MR, Hassani ME, Sharp PJ (2001) A PCR-based DNA marker for detection of mutant and normal alleles of the *Wx-D1* gene of wheat. Plant Breed 120:121–124

Somers DJ, Robert K, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109: 122–128

Sorrells ME, Rota ML, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Dong-Woog C, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Sreenivasulu N, Kavi Kishor PB, Varshney RK, Altschmied L (2002) Mining functional information from cereal genomes – the utility of expressed sequence tags (ESTs). Curr Sci 83:965–973

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney R, Perovic D, Grosse I, Graner A (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

Tang J, Gao L, Cao Y, Jia J (2006) Homologous analysis of SSR-ESTs and transferability of wheat SSR-EST markers across barley, rice and maize. Euphytica 151:87–93

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome 14:1812–1819

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Varshney RK, Beier U, Khlestkina EK, Kota R, Korzun V, Graner A, Börner A (2007a) Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency and applications for genome mapping and diversity studies. Theor Appl Genet 114:1105–1116 (doi 10.1007/s00122–007-0504–7)

Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Graner A, Sorrells ME (2005b) Genomic-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet 113:239–250

Varshney RK, Korzun V, Börner A (2004) Molecular maps in cereals: Methodology and progress. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, The Netherlands, p 35

Varshney RK, Marcel TC, Ramsay L, Russell J, Röder M, Stein N, Waugh R, Langridge P, Niks RE, Graner A (2007b) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet (DOI 10.1007/S00122-007-0503-7)

Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005c) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Wang HY, Liu DC, Yan ZH, Wei YM, Zheng YL (2005) Cytological characteristics of hybrid F2 population between *Triticum aestivum* L. and *T. durum* with reference to wheat breeding. J Appl Genet 46:365–369

Willsmore KL, Eckermann P, Varshney RK, Graner A, Langridge P, Pallotta M, Cheong J, Williams KJ (2006) New eSSR and gSSR markers added to Australian barley maps. Aust J Agric Res 57:953–959

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Cardle L, Ramsay L, Gibby M, Powell W (2003) Development of EST-SSRs from the alpine lady-fern, *Athyrium distentifolium*. Mol Eco Notes 3:287–290

Yu JK, La Rota M, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139